

COMMENT

Physics of Large Deviation

Shin-ichi Sasa

Department of Basic Science, The University of Tokyo, Tokyo 153-802, Japan

E-mail: sasa@jiro.c.u-tokyo.ac.jp

Abstract. A large deviation function mathematically characterizes the statistical property of atypical events. Recently, in non-equilibrium statistical mechanics, large deviation functions have been used to describe universal laws such as the fluctuation theorem. Despite such significance, large deviation functions have not been easily obtained in laboratory experiments. Thus, in order to understand the physical significance of large deviation functions, it is necessary to consider their experimental measurability in greater detail. This aspect of large deviation is discussed with the presentation of a future problem.

1. Introduction

When observing a time series of some fluctuating quantity at discrete times, one may find an atypical event that occurs during a time interval. In many cases, atypical events might be negligible because of the low frequency of their occurrence. However, there might be cases where such atypical events lead to a substantial effect, as in the cases of seismic events, economic activities, and evolution of species. In such cases, it is significant to determine the frequency of atypical events.

Atypical events are precisely formulated as follows. For mathematical simplicity, a time series is assumed to be described by a discrete-time Markov chain for a finite set of states. The transition probability from y to x is denoted by T_{xy} . The matrix T satisfies $\sum_x T_{xy} = 1$, and it is assumed to be irreducible. The initial state at time 0 is fixed as x_0 . The probability of a trajectory $\omega = (x_1, x_2, \dots, x_\tau)$ is expressed as

$$\mathcal{P}(\omega) = T_{x_\tau x_{\tau-1}} \cdots T_{x_1 x_0}. \quad (1)$$

Let J_{xy} be a quantity defined at the transition to x from y . The time average of J_{xy} during the time interval $[0, \tau]$ is expressed as

$$\mathcal{J}(\omega) = \frac{1}{\tau} \sum_{n=1}^{\tau} J_{x_n x_{n-1}}. \quad (2)$$

Here, the law of large numbers is assumed. That is, $\mathcal{J}(\omega)$ almost surely converges to a value J_* in the limit $\tau \rightarrow \infty$. When $\mathcal{J}(\omega)$ deviates from the typical value J_* , the time

series ω is considered to be atypical. The probability of the deviation is low for finite but large τ ; this probability is expressed as an exponential function of τ :

$$\text{Prob}(J) \simeq \exp(-\tau I(J)) \quad (3)$$

for large τ [1, 2, 3, 4]. This is called a large deviation property with a large deviation function $I(J)$. The frequency of atypical events is characterized by $I(J)$.

At this point, let us recall that the fluctuation theorem [5], which is a landmark in the recent development of non-equilibrium statistical mechanics, states the symmetry property of the large deviation function of the entropy production. The fluctuation theorem holds for a wide class of systems and is useful for deriving several non-equilibrium relations in a systematic manner. However, despite universal validity, its applications are still limited to the formal aspects of non-equilibrium systems.

It should be noted that large deviation functions are hardly measurable in experiments, except in the case of small fluctuations, because they characterize a small frequency of atypical events. For this reason, physicists have not encountered large deviation functions in experiments. If the term “physical quantity” is used to describe a quantity that can be measured in experiments, then a large deviation function cannot be considered a physical quantity. Such disadvantage in the experimental measurability of large deviation functions may be related to limited understanding of the physical significance of the fluctuation theorem. Thus, it would be a remarkable achievement if an experimental method for quickly obtaining $I(J)$ were to be devised. Recent development and future prospects of this achievement are briefly introduced in this Comment, as is implied by the title of this Comment, “Physics of Large Deviation.”

2. Basic idea

The key idea of the experimental determination of $I(J)$ involves a biased ensemble expressed as

$$\mathcal{P}^h(\omega) \equiv \frac{e^{h\tau\mathcal{J}(\omega)}\mathcal{P}(\omega)}{Z(h, \tau)}, \quad (4)$$

where the normalization constant $Z(h, \tau)$ is determined by

$$Z(h, \tau) = \sum_{\omega} e^{h\tau\mathcal{J}(\omega)}\mathcal{P}(\omega). \quad (5)$$

The quantity h in (4) is called a biasing field or a counting field. Here, the saddle-point calculation result for large τ is as follows:

$$\begin{aligned} Z(h, \tau) &\simeq \sum_J e^{\tau(hJ - I(J))}, \\ &\simeq e^{\tau \sup_J (hJ - I(J))}, \end{aligned} \quad (6)$$

which leads to the definition of $G(h)$ as

$$G(h) \equiv \sup_J [hJ - I(J)]. \quad (7)$$

This is the Legendre transform of $I(J)$. The inverse transformation is then expressed as

$$I(J) = \sup_h [hJ - G(h)]. \quad (8)$$

This expression indicates that the large deviation function $I(J)$ is equivalent to $G(h)$. $G(h)$ is called the scaled cumulant generating function, because the k -th order derivative of $G(h)$ at $h = 0$ is related to the k -th order cumulant of \mathcal{J} .

Now, suppose that the biased ensemble (4) can be generated experimentally. Let $\langle \rangle^h$ be the expectation with respect to the biased ensemble (4). Then, $\langle \mathcal{J} \rangle^h$ can be easily obtained in the experiments. Let $J_{\text{st}}^h = \lim_{\tau \rightarrow \infty} \langle J \rangle^h$. From (6), the following equation is derived.

$$J_{\text{st}}^h = \frac{dG(h)}{dh}. \quad (9)$$

The integration of this relation yields

$$G(h) = \int_0^h dh' J_{\text{st}}^{h'}. \quad (10)$$

That is, $G(h)$ (and $I(J)$ through (8)) is determined by the measurement of $J_{\text{st}}^{h'}$ for $0 \leq h' \leq h$. Thus, the problem can be solved by generating the biased ensemble (4) in the experiments.

It should be noted that the biased ensemble can be generated in numerical experiments by a cloning method. (See Ref. [6] for a recent related study in non-equilibrium statistical mechanics.) However, it cannot be performed in laboratory experiments. Experimental operations are those of adding an additional force or changing the temperature, which may be expressed by modifications of the transition matrix. Now, the problem becomes that of determining the modified transition matrix T_{xy}^h that generates the biased ensemble $\mathcal{P}^h(\omega)$ in the large τ limit.

3. Variational principle

The transition matrix T^h is characterized by a variational principle. Its mathematical derivation is described below. A set of transition matrices compatible with T_{xy} is denoted by \mathcal{V}_T . (For $R_{xy} \in \mathcal{V}_T$, $R_{xy} > 0$ only when $T_{xy} > 0$.) Let $\mathcal{R}(\omega)$ be the path probability generated by a transition matrix R in \mathcal{V}_T . The application of Jensen's inequality to the trivial identity

$$G(h) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \sum_{\omega} \mathcal{R}(\omega) \frac{e^{h\tau \mathcal{J}(\omega)} \mathcal{P}(\omega)}{\mathcal{R}(\omega)} \quad (11)$$

leads to

$$G(h) \geq \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{\omega} \mathcal{R}(\omega) \left[h\tau \mathcal{J}(\omega) - \log \frac{\mathcal{R}(\omega)}{\mathcal{P}(\omega)} \right]. \quad (12)$$

Owing to the law of large numbers, this is rewritten as

$$G(h) \geq \sum_{xy} R_{xy} p_y^R \left[h J_{xy} - \log \frac{R_{xy}}{T_{xy}} \right], \quad (13)$$

where p^R is the stationary distribution of the transition matrix R . (That is, $\sum_y R_{xy} p_y^R = p_x^R$.)

On the other hand, (4) and (6) lead to

$$G(h) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left[h \tau \mathcal{J}(\omega) - \log \frac{\mathcal{P}^h(\omega)}{\mathcal{P}(\omega)} \right]. \quad (14)$$

Multiplying the both-hand sides by $\mathcal{P}^h(\omega)$ and taking the summation over all histories ω , one has

$$G(h) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{\omega} \mathcal{P}^h(\omega) \left[h \tau \mathcal{J}(\omega) - \log \frac{\mathcal{P}^h(\omega)}{\mathcal{P}(\omega)} \right]. \quad (15)$$

Again, owing to the law of large numbers, (15) is rewritten as

$$G(h) = \sum_{xy} T_{xy}^h p_y^{T^h} \left[h J_{xy} - \log \frac{T_{xy}^h}{T_{xy}} \right]. \quad (16)$$

From (13) and (16), the following variational formula is obtained:

$$G(h) = \max_{R \in \mathcal{V}_T} \Phi(h, R) \quad (17)$$

and

$$T^h = \operatorname{Argmax}_{R \in \mathcal{V}_T} \Phi(h, R), \quad (18)$$

where the function $\Phi(h, R)$ has been defined as

$$\Phi(h, R) \equiv \sum_{xy} R_{xy} p_y^R \left[h J_{xy} - \log \frac{R_{xy}}{T_{xy}} \right]. \quad (19)$$

Formula (17) is a well-known formula in large deviation theory. (See, for example, page 81 in Ref. [1] or page 284 in Ref. [2]. See also Ref. [7], which reports the formula.)

Furthermore, from the Perron-Frobenius theory for irreducible matrices, it is found that there exist a positive vector ϕ_x^* and a positive constant λ such that

$$\sum_x \phi_x^* e^{h J_{xy}} T_{xy} = \lambda \phi_y^*, \quad (20)$$

where (5) and (7) lead to $G(h) = \log \lambda$. Then, with the definition

$$R_{xy}^{h, \phi} \equiv \phi_x T_{xy} e^{h J_{xy}} \phi_y^{-1} \mu \quad (21)$$

for any positive vector ϕ , where the constant μ is determined as the normalization condition $\sum_x R_{xy}^{h, \phi} = 1$, it is confirmed that $T^h = R^{h, \phi^*}$. (Substitute $T^h = R^{h, \phi^*}$ into (16).) The result indicates that

$$G(h) = \max_{\phi > 0} \Phi(h, R^{h, \phi}) \quad (22)$$

and

$$\phi^* = \underset{\phi > 0}{\text{Argmax}} \Phi(h, R^{h, \phi}). \quad (23)$$

That is, the set of variational parameters \mathcal{V}_T is reduced to a set of positive vectors.

In this manner, the transition matrix that generates the biased ensemble is characterized by the variational principle. The best transition matrix to optimize the function $\Phi(h, R)$ can be obtained by considering modifications to the system. The formulas for continuous-time Markov jump processes and Langevin systems can be directly obtained by considering an appropriate limit for (22) with (19). The resulting formulas are equivalent to those reported in Refs. [10, 11]. It should be noted that the formulas are considered to be an extension of those derived for systems with the detailed balance property [8, 9].

Now, the main message of Refs. [10, 11] is that *the variational parameter ϕ_x corresponds to a potential function for physical models* that are described by continuous time Markov jump processes and Langevin systems. That is, for these cases, the optimization can be considered in laboratory experiments. This indicates that the biased ensemble is generated experimentally so that the large deviation function can be obtained quickly in the experiments. The experimental determination of the large deviation function for a single Brownian particle under a non-equilibrium condition was demonstrated in Refs. [10, 11].

4. Future problem

The result appears to be very promising. However, the complication in the result can be observed immediately. In the case of a single particle, only the external potential needs to be controlled in order to obtain a biased ensemble; however, the degrees of freedom of dynamical evolution rules increase exponentially with the particle number. For example, when a system consisting of two particles is considered, one needs to modify the interaction potential between the particles in addition to the modification of the one-body potential. In this regard, (23) is formally correct; however, it may not be applicable to many-body systems.

A one-dimensional lattice gas model consisting of N sites in contact with particle reservoirs at two ends is studied so as to investigate the problem more concretely. The state of the system is denoted by $\sigma = (\sigma_j)_{j=1}^N$, where $\sigma_j \in \{0, 1\}$, and the transition rate in the Markov jump process (or the transition matrix in the Markov chain) is expressed as a $2^N \times 2^N$ matrix. Now, by using (22), the optimization problem is formulated in the set of 2^N dimensional positive vectors. Each positive vector ϕ corresponds to a potential V as a function of σ . The potential can be formally expanded in the form

$$V(\sigma) = \sum_i V_i^{(1)} \sigma_i + \sum_{ij} V_{ij}^{(2)} \sigma_i \sigma_j + \cdots, \quad (24)$$

where, in principle, k -body, N -range interactions $V^{(k)}$ are included. The space of the variational parameters is too large to be controlled.

Let us assume that experimentalists can control only a one-body potential for a macroscopic system; this assumption corresponds to the case where only the first term in the expanded form (24) can be considered for a system with large N . An optimistic consideration that the description of macroscopic behaviors does not require the complete information of microscopic details of the system leads one to expect that the control of a one-body potential is sufficient to solve the optimization problem. On the other hand, cautious researchers would not have such an ill-founded expectation. In any case, the question may be mathematically stated as follows: *Can the optimization problem with a 2^N -dimensional vector as a variational parameter be effectively reduced to an optimization problem with an N -dimensional vector as a variational parameter in the macroscopic limit?*

Obviously, such a reduction is not applicable to all systems. Then, an interesting future problem is to obtain a condition under which the reduction indeed occurs. This problem may be related to the manner in which microscopic descriptions are connected with macroscopic descriptions in non-equilibrium systems. Hopefully, studying this future problem will provide a new direction to non-equilibrium statistical mechanics.

Acknowledgments

The author thanks T. Nemoto for the collaborative work on “Physics of Large Deviation.” The present study was supported by KAKENHI Nos. 22340109 and 23654130.

References

- [1] Dembo A and Zeitouni O 1998 *Large deviations techniques and applications*, (Springer, New York).
- [2] Ellis R S 1985 *Entropy, Large deviations, and Statistical Mechanics*, (Springer, New York).
- [3] Oono Y 1989 Prog. Theor. Phys. Suppl. **99** 165
- [4] Touchette H 2009 Phys. Rep. **478** 1
- [5] Evans D J, Cohen E G D, and Morriss G P 1993 Phys. Rev. Lett. **71** 2401
- [6] Giardinà C, Kurchan J, Lecomte V, and Tailleur J 2011 J. Stat. Phys. **145** 787
- [7] Andrieux D 2012 arXiv:1201.1232v1
- [8] Garrahan J P, Jack R L, Lecomte V, Pitard E, van Duijvendijk K, and van Wijland F 2009 J. Phys. A, Math. Theor. **42** 075007
- [9] Jack R and Sollich P 2010 Prog. Theor. Phys. Supp. **184** 304
- [10] Nemoto T and Sasa S 2011 Phys. Rev. E **83** 030105(R)
- [11] Nemoto T and Sasa S 2011 Phys. Rev. E **84** 061113